E1 244: Detection & Estimation Theory **Jan-May 2017**

Lecture 10 — February 7

Lecturer: Parimal parag *Scribe: P Shiva Kumar & K Vikas Bharadwaj*

SIGNAL DETECTION IN DISCRETE TIME

10.1 Recap

We have studied optimum detector structures for coherent symbols in discrete time in presence of i.i.d gaussian noise and i.i.d laplacian noise. We also studied locally optimum detection of coherent symbols in i.i.d gaussian and laplacian noises where we assumed that the form of signal detected is known but not its amplitude.

10.2 Deterministic signal detection in gaussian dependent noise

The basic problem that we are dealing is the following hypothesis testing problem:

$$H_0 : \bar{Y} = \bar{S}_0 + \bar{N}$$

$$H_1 : \bar{Y} = \bar{S}_1 + \bar{N}$$

where

$$\bar{Y} = [Y_1 \quad Y_2 \dots Y_n]^T$$

is a vector in the observation space, $\mathcal{R}^n$,

$$\bar{S}_0 = [s_{01} \quad s_{02} \dots s_{0n}]^T$$

$$\bar{S}_1 = [s_{11} \quad s_{12} \dots s_{1n}]^T$$

are the two deterministic and known signals, in discrete time and

$$\bar{N} = [N_1 \quad N_2 \dots N_n]^T$$

is the vector of noise samples that are added to the signals and is gaussian with zero mean and covariance matrix Σ_N, i.e., $\mathbb{E}\{\bar{N}\bar{N}^T\} = \Sigma_N$

The likelihood ratio corresponding to an observation $\bar{y}$ is given by

$$L(\bar{y}) = \frac{\mathbb{E}_1[p_N(\bar{y} - \bar{S}_1)]}{\mathbb{E}_0[p_N(\bar{y} - \bar{S}_0)]}$$

Remark 1 $\mathbb{E}_i$ *denotes expectation over* $\bar{S}_i$ *for* $i = \{0, 1\}$ *and* $p_{\bar{N}}(x)$ *denotes the probability density function of* $\bar{N}$ *at a vector* x.

Since the signals $\bar{S_0}$ and $\bar{S_1}$ are deterministic, the likelihood ratio simplifies to

$$L(\bar{y}) = \frac{p_{\bar{N}}(\bar{y} - \bar{S_1})}{p_{\bar{N}}(\bar{y} - \bar{S_0})} \tag{10.1}$$

For a gaussian vector $\bar{X}$ with mean $\bar{\mu}$ and covariance matrix Σ_N, its probability density function is given by

$$p_{\bar{X}}(x) = \frac{1}{(2\pi)^{n/2} det(\Sigma_N))^{1/2}} exp[-\frac{1}{2}(\bar{X} - \bar{\mu})^T {\Sigma_N}^{-1} (\bar{X} - \bar{\mu})]$$

Then the likelihood ratio in (10.1) becomes

$$\begin{aligned}
L(\bar{y}) &= \frac{\frac{1}{(2\pi)^{n/2} det(\Sigma_N))^{1/2}} exp[-\frac{1}{2}(\bar{y} - \bar{S_1})^T {\Sigma_N}^{-1} (\bar{y} - \bar{S_1})]}{\frac{1}{(2\pi)^{n/2} det(\Sigma_N))^{1/2}} exp[-\frac{1}{2}(\bar{y} - \bar{S_0})^T {\Sigma_N}^{-1} (\bar{y} - \bar{S_0})]} \\
&= exp[\frac{1}{2}(\bar{y} - \bar{S_0})^T {\Sigma_N}^{-1} (\bar{y} - \bar{S_0}) - \frac{1}{2}(\bar{y} - \bar{S_1})^T {\Sigma_N}^{-1} (\bar{y} - \bar{S_1})] \\
&= exp[\bar{S_1}^T {\Sigma_N}^{-1} \bar{y} - \bar{S_0}^T {\Sigma_N}^{-1} \bar{y} - \frac{1}{2}\bar{S_1}^T {\Sigma_N}^{-1} \bar{S_1} + \frac{1}{2}\bar{S_0}^T {\Sigma_N}^{-1} \bar{S_0}] \\
&= exp[(\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1} (\bar{y} - \frac{\bar{S_0} + \bar{S_1}}{2})]
\end{aligned}$$

Upon taking natural logarithm, we have

$$\ln(L(\bar{y})) = (\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1} (\bar{y} - \frac{\bar{S_0} + \bar{S_1}}{2})$$

$$\Rightarrow \quad \ln(L(\bar{y})) = (\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1} \bar{y} - \frac{(\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1} (\bar{S_0} + \bar{S_1})}{2}$$

Now by defining $\tilde{S}^T := (\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1}$, it can be seen that $T(\bar{Y}) \triangleq \tilde{S}^T \bar{Y}$ is a linear transformation of the gaussian random vector $\bar{Y}$ and hence the decision rule is given by

$$\tilde{\delta}_0(\bar{y}) = \begin{cases} 1 & if \quad \tilde{S}^T \bar{y} > \tau'' \\ \gamma & if \quad \tilde{S}^T \bar{y} = \tau'' \\ 0 & if \quad \tilde{S}^T \bar{y} < \tau'' \end{cases} \tag{10.2}$$

where

$$\tau'' = \ln(\tau) + \frac{(\bar{S_1} - \bar{S_0})^T \Sigma_N^{-1} (\bar{S_0} + \bar{S_1})}{2}$$

The structure of the detector is similar to that of a correlation detector is shown below.

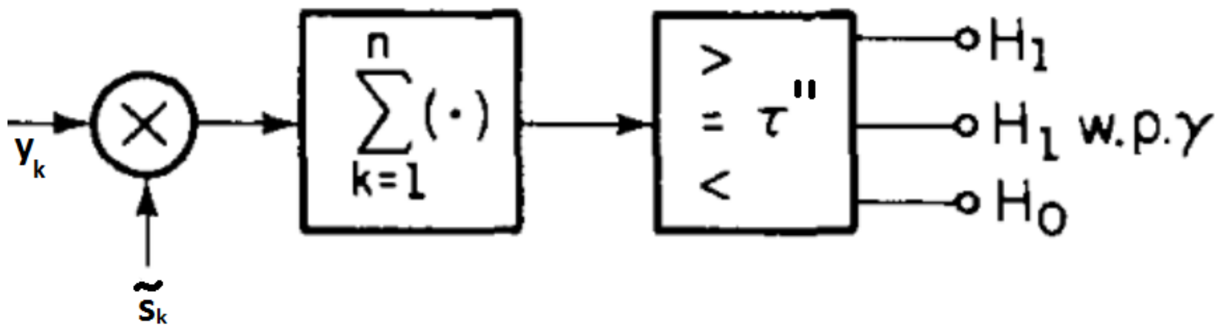

Figure 10.1: Implementation of detector in (10.2)

A property of multivariate gaussian distributions is that their linear transformations are also gaussian. Hence $T(\bar{Y})$ is a gaussian random variable and thus, we can characterize its distribution under H_0 and H_1 completely by finding its mean and variance under the two hypotheses.

Under H_j the mean of $T(\bar{Y})$ is given by

$$\begin{aligned}\mathbb{E}\{T(\bar{Y})|H_j\} &= \mathbb{E}\{\tilde{S}^T\bar{Y}|H_j\}\\ &= \tilde{S}^T\mathbb{E}\{\bar{Y}|H_j\}\\ &= \tilde{S}^T\bar{S}_j \triangleq \tilde{\mu}_j\end{aligned}$$

Similarly, the variance of $T(\bar{Y})$ under H_j is

$$\begin{aligned}Var(T(\bar{Y})|H_j) &= \mathbb{E}\{(\tilde{S}^T(\bar{Y}-\bar{S}_j))^2|H_j\}\\ &= \mathbb{E}\{(\tilde{S}^T\bar{N})^2|H_j\}\\ &= \mathbb{E}\{(\tilde{S}^T\bar{N})^2\}\\ &= \mathbb{E}\{\tilde{S}^T\bar{N}\bar{N}^T\tilde{S}\}\\ &= \tilde{S}^T\varSigma_N\tilde{S}\\ &= (\bar{S}_1-\bar{S}_0)^T\varSigma_N^{-1}(\bar{S}_1-\bar{S}_0) =: d^2\end{aligned}$$

Note that the variance of $T(\bar{Y})$ is independent of the hypotheses. Also note that positive definiteness of $\varSigma_N$ imples positive definiteness of $\varSigma_N^{-1}$ and thus that $d^2 > 0$ unless $\bar{S}_0 = \bar{S}_1$.

From the analysis above, we see that $T(\bar{Y}) \sim \mathcal{N}(\tilde{S}^T\bar{S}_j, d^2)$ under H_j for $j = 0, 1$ and Γ_1 can be formulated as,

$$\Gamma_1 = \{\bar{y} : T(\bar{y}) > \tau''\}$$

See that the randomized partition of Γ in eqn(10.2) is irrelevant because of the continuity of $T(\bar{y})$. The probability of deciding H_1 under H_j is thus given by

$$\begin{aligned} p_j(\Gamma_1) &= \int_{\Gamma_1} \frac{1}{\sqrt{2\pi d^2}} exp(-\frac{(\bar{y} - \tilde{S}^T \bar{S}_j)^2}{2d^2}) d\bar{y} \\ &= 1 - \mathbf{\Phi}(\frac{\tau'' - \tilde{S}^T \bar{S}_j}{d}) \end{aligned}$$

where $\mathbf{\Phi}(.)$ is the standard normal distribution function. Upon substituting τ'' in $p_j(\Gamma_1)$, we have

$$p_j(\Gamma_1) = \begin{cases} 1 - \mathbf{\Phi}(\frac{\ln \tau}{d} + \frac{d}{2}) & if \quad j = 0 \\ 1 - \mathbf{\Phi}(\frac{\ln \tau}{d} - \frac{d}{2}) & if \quad j = 1 \end{cases}$$

Then we can have expressions for detection probability and false alarm probability.

Example 10.2.1 *For an α-level Neyman Pearson testing we set,* $\alpha = p_F(\tilde{\delta}_0) = p_0(\Gamma_1)$

$$i.e., \alpha = 1 - \mathbf{\Phi}(\tfrac{\tau'' - \tilde{\mu}_0}{d})$$

$$\Rightarrow \tau'' = d\mathbf{\Phi}^{-1}(1-\alpha) + \tilde{\mu}_0$$

now, the corresponding detection probablity becomes

$$\begin{aligned} p_1(\Gamma_1) &= 1 - \mathbf{\Phi}(\frac{\tau'' - \tilde{\mu}_1}{d}) \\ &= 1 = \mathbf{\Phi}(\frac{d\mathbf{\Phi}^{-1}(1-\alpha) + \tilde{\mu}_0 - \tilde{\mu}_1}{d}) \\ &= 1 - \mathbf{\Phi}(\mathbf{\Phi}^{-1}(1-\alpha) + \frac{\tilde{\mu}_0 - \tilde{\mu}_1}{d}) \end{aligned}$$

$$\begin{aligned} \frac{\tilde{\mu}_0 - \tilde{\mu}_1}{d} &= \frac{\tilde{S}^T(\bar{S}_0 - \bar{S}_1)}{d} \\ &= \frac{(\bar{S}_1 - \bar{S}_0)^T \Sigma_N^{-1} (\bar{S}_0 - \bar{S}_1)}{d} \\ &= \frac{-d^2}{d} \\ &= -d \end{aligned}$$

$$\Rightarrow p_1(\Gamma_1) = p_D(\tilde{\delta}_{NP}) = 1 - \mathbf{\Phi}(\mathbf{\Phi}^{-1}(1-\alpha) - d)$$

10.3 Interpretation of $\mathbf{d^2}$:

In view of the discussion above we see that the performance of optimum detection of deterministic signals in gaussian noise improves monotonically with increasing d.This quantity (or more properly its square)can be interpretaed as a measure of signal-to-noise ratio. To see this, consider, without loss of generality, signals $\bar{S}_0 = \bar{0}, \bar{S}_1 = \bar{s}$. The noise samples can be independent or dependent.

10.3.1 I.I.D noise

$\bar{N}$ is a zero mean gaussian random vector with covariance matrix $\Sigma_N = \sigma^2 I$, where I denotes the $n \times n$ identity matrix.
Therefore, $\Sigma_N^{-1} = \frac{1}{\sigma^2} I$. Hence

$$\begin{aligned} d^2 &= (\bar{S}_1 - \bar{S}_0)^T \Sigma_N^{-1} (\bar{S}_1 - \bar{S}_0) \\ &= \frac{(\bar{S}_1 - \bar{S}_0)^T (\bar{S}_1 - \bar{S}_0)}{\sigma^2} \\ &= \sum_{k=1}^{n} \frac{(s_1 - s_0)_k^2}{\sigma^2} \\ &= \frac{1}{\sigma^2} \sum_{k=1}^{n} (s_1 - s_0)_k^2 \end{aligned}$$

upon substituting $\bar{S}_0 = \bar{0}, \bar{S}_1 = \bar{s}$,we get

$$\begin{aligned} d^2 &= \frac{1}{\sigma^2} \sum_{k=1}^{n} s_k^2 \\ &= n \frac{\bar{s^2}}{\sigma^2} \end{aligned}$$

here $\bar{s^2}$: is the Average signal power and
σ^2: is the Average noise power, mathematically formulated as,

$\bar{s^2} \triangleq \frac{1}{n} \sum_{k=1}^{n} s_k^2$

$\sigma^2 \triangleq \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} N_k^2$
Therefore d^2 here is given by the average signal-to-noise ratio times the number of samples. Thus performance is enhanced by increasing either of these quantities, and as either of the two increases without bound perfect performance can result.

10.3.2 Non-i.i.d noise

A similar interpretation can be given to d^2 in the non-i.i.d. case where $\bar{N}$ is a zero mean gaussian random vector with covariance matrix Σ_N.We can write the quantity $\sum_{k=1}^{n} \tilde{S}_k y_k$ as

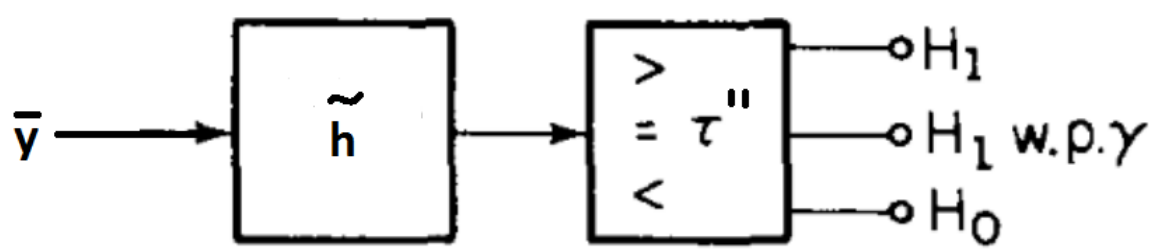

Figure 10.2: Implementation of detector in the form of an LTI filter

the input at time n of a linear time-invariant filter with impulse response

$$\tilde{h}_k = \begin{cases} \tilde{S}_{n-k} & for \quad 0 \leq k \leq n-1 \\ 0 & otherwise \end{cases}$$

If the input to this filter is $\bar{s} = (s_1 \dots s_n)$,then the output at time n would be

$$\begin{aligned} \sum_{k=1}^{n} \tilde{S}_k s_k &= \bar{s}^T \Sigma_N^{-1} \bar{s} \\ &= d^2 \end{aligned}$$

Thus the output power at the sampling time due to signal only is d^4. On the other hand, if the noise only were put into this filter, the output at time n would be $\sum_{k=1}^{n} \tilde{S}_k N_k$ a random quantity with variance

$$\begin{aligned} \mathbb{E}(\sum_{k=1}^{n} \tilde{S}_k N_k)^2 &= \mathbb{E}(\bar{\tilde{S}}^T \bar{N})^2 \\ &= \bar{\tilde{S}}^T \Sigma_N \bar{\tilde{S}} \\ &= \bar{s}^T \Sigma_N^{-1} \Sigma_N \Sigma_N^{-1} \bar{s} \\ &= \bar{s}^T \Sigma_N^{-1} \bar{s} \\ &= d^2 \end{aligned}$$

The ratio of signal power to noise variance at the output of filter at each sampling instant is

$$\begin{aligned} \frac{(\sum_{k=1}^{n} \tilde{S}_k s_k)^2}{\mathbb{E}\{(\sum_{k=1}^{n} \tilde{S}_k N_k)^2\}} &= \frac{d^4}{d^2} \\ &= d^2 \end{aligned}$$

Thus the quantity d^2 in the general case is the signal-to-noise power ratio at the output of the filter used for optimum detection at the sampling time n.It is intuitively reasonable that the higher this output SNR is, the better the signal can be detected by comparing the sampled output to a threshold, and this intuition is borne out by the monotonicity of detection of performance as a function of d^2 shown above.
The quantity d^2 also has another interpretation for the i.i.d. case with general signals $\bar{S}_1, \bar{S}_0$

$$\bar{N} \sim i.i.d \quad and \quad \bar{N} \sim \mathcal{N}(\bar{0}, \sigma^2 I)$$

in this case, we can write $d^2 = \frac{\|\bar{S}_1 - \bar{S}_0\|^2}{\sigma^2}$, where $\| \bar{S}_1 - \bar{S}_0 \|$ denotes the Euclidean distance between the signal vectors $\bar{S}_0$ and $\bar{S}_1$ given by

$$\| \bar{S}_1 - \bar{S}_0 \|_2 = [\sum_{k=1}^{n}(s_{1k} - s_{0k})^2]^{1/2}$$

Thus the farther apart the signal vectors are, the better performance can be achieved. A similar interpretation can be made for non-i.i.d. noise case also.

10.4 Reduction to the i.i.d. noise case:

Since Σ_N is positive definite, it can be written as

$$\Sigma_N = CC^T$$

where C is an $n \times n$ invertible lower triangular matrix and also above form of Σ_N is called the *Cholesky decomposition* of Σ_N and there are several standard algorithms for finding C from Σ_N. We can write,

$$\begin{aligned}\Sigma_N^{-1} &= (CC^T)^{-1} \\ &= (C^T)^{-1}C^{-1} \\ &= (C^{-1})^T C^{-1}\end{aligned}$$

Let $D = C^{-1}$, then we have $\quad \Sigma_N^{-1} = D^T D$

Defining new variables,

$$\hat{S}_j := D\bar{S}_j$$

$$\hat{N} := D\bar{N}$$

$$\hat{Y} := D\bar{Y}$$

The appropriate linear transformation of $\hat{Y}$ is $T(\hat{Y})$. However $\bar{N}$ is the i.i.d. and $\sim \mathcal{N}(\bar{0}, I)$

$$\begin{aligned}\mathbb{E}(\hat{N}) &= \mathbb{E}(D\bar{N}) \\ &= D\mathbb{E}(\bar{N}) \\ &= \bar{0}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(\hat{N}\hat{N}^T) &= \mathbb{E}(D\bar{N}\bar{N}^T D^T) \\ &= D\mathbb{E}(\bar{N}\bar{N}^T)D^T \\ &= D\varSigma_N D^T \\ &= DCC^T D^T \\ &= I\end{aligned}$$

Therefore, the optimum detection statistic is

$$\begin{aligned}T(\hat{Y}) &= (\bar{S}_1 - \bar{S}_0)^T \varSigma_N^{-1} \bar{Y} \\ &= (\bar{S}_1 - \bar{S}_0)^T D^T D\bar{Y} \\ &= (D(\bar{S}_1 - \bar{S}_0))^T D\bar{Y} \\ &= (\hat{S}_1 - \hat{S}_0)^T \hat{Y}\end{aligned}$$

The interesting thing about this particular transformation is that the lower triangularity of C implies that C^{-1} is also lower triangular.

It can be seen that implementation of $T(\hat{Y})$ as an LTI filter is a causal operation. Since the noise in the output of this filter is white this filter is also known as *whitening filter*.

As a final comment we note that the signal-to-noise ratio

$$d^2 = (\bar{S}_1 - \bar{S}_0)^T \varSigma_N^{-1} (\bar{S}_1 - \bar{S}_0)$$

can be written interms of the transformed signal pairs

$$d^2 = \| \bar{S}_1 - \bar{S}_0 \|_2^{\,2} = \| \hat{S}_1 - \hat{S}_0 \|_2^{\,2} \tag{10.3}$$

Thus the performance of coherent detection in dependent noise depends on how far apart the signals are when transformed to a coordinate system in which the noise components are i.i.d.. It should be noted that all signals pairs in (10.3) are the same distance apart because they are all representations of the same pair of vectors in different coordinate systems that are simple rotations of one another.

10.5 Signal design

The performance of optimum coherent detection in gaussian noise is improved by incresing the quantity

$$d^2 \triangleq (\bar{S}_1 - \bar{S}_0)^T \varSigma_N^{-1} (\bar{S}_1 - \bar{S}_0)$$

In many of the applications in which coherent detection arises, there is often some flexibility in the choice of the signals $\bar{S}_0 \quad and \quad \bar{S}_1$. In such situations it is reasonable to choose these signals to maximize d^2. However the signals are usually constrained by their total power P.Let us consider the case in which

$$\bar{S}_0 = \bar{0} \quad and \quad \bar{S}_1 = \bar{S}$$

Therefore the problem statement can be posed as,

$$\max_{\bar{S}} \quad (d^2) = \max_{\bar{S}} \quad (\bar{S}^T \varSigma_N^{-1} \bar{S}) \qquad s.t. \quad \| \bar{S} \|_2{}^2 \leq P$$

Since $\varSigma_N$ is an $n \times n$ symmetric positive definite matrix,it has several structural properties that can be examined to give some insight into the structure of the optimum detection system.The eigen values $\lambda_1, \ldots \lambda_n$ and corresponding eigen vector $\bar{V}_1, \ldots \bar{V}_k$ of the $n \times n$ matrix $\varSigma_N$ are the solutions to the equation $\varSigma_N \bar{V}_k = \lambda_k \bar{V}_k$. Since $\varSigma_N$ in our case is symmetric and positive definite, all of its eigen values are real and positive and its eigen vectors can be choosen to be orthonormal.i.e.,

$$\bar{V}_k^T v_l = \begin{cases} o & if \quad k \neq l \\ 1 & if \quad k = l \quad l, k = 1, \ldots n. \end{cases}$$

with this choice of eigen vector we can write $\varSigma_N$ as

$$\varSigma_N = \sum_{k=1}^{n} \lambda_k \bar{V}_k \bar{V}_k^T \tag{10.4}$$

(10.4) is called the *Spectral decomposition* of $\varSigma_N$.And hence

$$\varSigma_N^{-1} = \sum_{k=1}^{n} \lambda_k{}^{-1} \bar{V}_k \bar{V}_k^T$$

So for any vector $\bar{x} \in \mathcal{R}^n$, we have

$$\begin{aligned} \bar{x}^T \varSigma_N^{-1} \bar{x} &= \sum_{k=1}^{n} \lambda_k^{-1} \bar{x}^T \bar{V}_k \bar{V}_k^T \bar{x} \\ &\leq \lambda_{min}^{-1} \sum_{k=1}^{n} \bar{x}^T \bar{V}_k \bar{V}_k^T \bar{x} \\ &= \lambda_{min}^{-1} \bar{x}^T (\sum_{k=1}^{n} \bar{V}_k \bar{V}_k^T) \bar{x} \\ &= \lambda_{min}^{-1} \bar{x}^T \bar{x} \\ &= \lambda_{min}^{-1} \| \bar{x} \|_2{}^2 \end{aligned}$$

where $\lambda_{min} = min\{\lambda_1 \dots \lambda_n\}$.

$$\therefore \bar{x}^T \Sigma_N^{-1} \bar{x} \leq \lambda_{min}^{-1} \| \bar{x} \|_2{}^2 \tag{10.5}$$

Note that we can have equality in (10.5) if and only if $\bar{x}$ is proportional to an eigen vector corresponding to the eigen value λ_{min}.

From the above we see that, for fixed $\| \bar{S} \|$, the best way to choose the signal $\bar{S}$ is to be along an eigen vector corresponding to the minimum eigen value of Σ_N i.e.,$\bar{S} = c\bar{V}_k \quad where \lambda_k = \lambda_{min}$

$$\Rightarrow \| \bar{S} \|_2{}^2 = c^2 \| \bar{V}_k \|_2{}^2 \Rightarrow c = \frac{\| \bar{S} \|_2}{\| \bar{V}_k \|_2}$$

Therefore,

$$\begin{aligned} \bar{S} &= \| \bar{S} \|_2 \frac{\bar{V}_k}{\| \bar{V}_k \|_2} \\ &= \| \bar{S} \|_2 \frac{\bar{V}_{min}}{\| \bar{V}_{min} \|_2} \end{aligned}$$

and

$$\begin{aligned} d^2 &= \bar{S}^T \Sigma_N^{-1} \bar{S} \\ &= \frac{\| \bar{S} \|_2 \bar{V}_k{}^T \Sigma_N^{-1} \bar{V}_k \| \bar{S} \|_2}{\| \bar{V}_k \|_2{}^2} \\ &= \| \bar{S} \|_2{}^2 \frac{\bar{V}_k{}^T \Sigma_N^{-1} \bar{V}_k}{\| \bar{V}_k \|_2{}^2} \\ &= \frac{\| \bar{S} \|_2{}^2}{\lambda_{min}} \end{aligned}$$

$$\therefore d^2 = \frac{\| \bar{S} \|_2{}^2}{\lambda_{min}} \tag{10.6}$$

Once we have chosen the direction of the signal $\bar{S}$, we can further optimize performance by maximizing $\| \bar{S} \|_2{}^2$. Obviously, this quantity can be arbitrary large if we put no constraint on the signal. But in our case it is limited by P. Hence maximizing $\| \bar{S} \|_2{}^2$ will result in maximizing d^2.
$\max \| \bar{S} \|_2{}^2 = P$.
From (10.6)

$$\max_{\bar{S}} \quad (d^2) = \frac{P}{\lambda_{min}}$$

. And the optimum signals are

$$\bar{S}_0 = \bar{0} \quad and \quad \bar{S}_1 = \sqrt{P} \frac{\bar{V}_{min}}{\| \bar{V}_{min} \|_2}$$

.